

2. Kant's project in the *Critique of Pure Reason*

The main topic of the *Critique of Pure Reason* is the possibility of metaphysics, understood in a specific way. Kant defines metaphysics in terms of "the cognitions after which reason might strive independently of all experience," and his goal in the book is to reach a "decision about the possibility or impossibility of a metaphysics in general, and the determination of its sources, as well as its extent and boundaries, all, however, from principles" (Axii. See also Bxiv; and 4:255–257). Thus metaphysics for Kant concerns a priori knowledge, or knowledge whose justification does not depend on experience; and he associates a priori knowledge with reason. The project of the *Critique* is to examine whether, how, and to what extent human reason is capable of a priori knowledge.

2.1 The crisis of the Enlightenment

To understand the project of the *Critique* better, let us consider the historical and intellectual context in which it was written.^[5] Kant wrote the *Critique* toward the end of the Enlightenment, which was then in a state of crisis. Hindsight enables us to see that the 1780's was a transitional decade in which the cultural balance shifted decisively away from the Enlightenment toward Romanticism, but Kant did not have the benefit of such hindsight.

The Enlightenment was a reaction to the rise and successes of modern science in the sixteenth and seventeenth centuries. The spectacular achievements of Newton in particular engendered widespread confidence and optimism about the power of human reason to control nature and to improve human life. One effect of this new confidence in reason was that traditional authorities were increasingly questioned. Why should we need political or religious authorities to tell us how to live or what to believe, if each of us has the capacity to figure these things out for ourselves? Kant expresses this Enlightenment commitment to the sovereignty of reason in the *Critique*:

Our age is the age of criticism, to which everything must submit. Religion through its holiness and legislation through its majesty commonly seek to exempt themselves from it. But in this way they excite a just suspicion against themselves, and cannot lay claim to that unfeigned respect that reason grants only to that which has been able to withstand its free and public examination. (Axi)

Enlightenment is about thinking for oneself rather than letting others think for you, according to *What is Enlightenment?* (8:35). In this essay, Kant also expresses the Enlightenment faith in the inevitability of progress. A few independent thinkers will gradually inspire a broader cultural movement, which ultimately will lead to

greater freedom of action and governmental reform. A culture of enlightenment is “almost inevitable” if only there is “freedom to make public use of one’s reason in all matters” (8:36).

The problem is that to some it seemed unclear whether progress would in fact ensue if reason enjoyed full sovereignty over traditional authorities; or whether unaided reasoning would instead lead straight to materialism, fatalism, atheism, skepticism (Bxxxiv), or even libertinism and authoritarianism (8:146). The Enlightenment commitment to the sovereignty of reason was tied to the expectation that it would not lead to any of these consequences but instead would support certain key beliefs that tradition had always sanctioned. Crucially, these included belief in God, the soul, freedom, and the compatibility of science with morality and religion. Although a few intellectuals rejected some or all of these beliefs, the general spirit of the Enlightenment was not so radical. The Enlightenment was about replacing traditional authorities with the authority of individual human reason, but it was not about overturning traditional moral and religious beliefs.

Yet the original inspiration for the Enlightenment was the new physics, which was mechanistic. If nature is entirely governed by mechanistic, causal laws, then it may seem that there is no room for freedom, a soul, or anything but matter in motion. This threatened the traditional view that morality requires freedom. We must be free in order to choose what is right over what is wrong, because otherwise we cannot be held responsible. It also threatened the traditional religious belief in a soul that can survive death or be resurrected in an afterlife. So modern science, the pride of the Enlightenment, the source of its optimism about the powers of human reason, threatened to undermine traditional moral and religious beliefs that free rational thought was expected to support. This was the main intellectual crisis of the Enlightenment.

The *Critique of Pure Reason* is Kant’s response to this crisis. Its main topic is metaphysics because, for Kant, metaphysics is the domain of reason – it is “the inventory of all we possess through pure reason, ordered systematically” (Axx) – and the authority of reason was in question. Kant’s main goal is to show that a critique of reason by reason itself, unaided and unrestrained by traditional authorities, establishes a secure and consistent basis for both Newtonian science and traditional morality and religion. In other words, free rational inquiry adequately supports all of these essential human interests and shows them to be mutually consistent. So reason deserves the sovereignty attributed to it by the Enlightenment.

2.2 Kant’s Copernican revolution in philosophy

To see how Kant attempts to achieve this goal in the *Critique*, it helps to reflect on his grounds for rejecting the Platonism of the *Inaugural Dissertation*. The *Inaugural Dissertation* also tries to

reconcile Newtonian science with traditional morality and religion in a way, but its strategy is different from that of the *Critique*. According to the *Inaugural Dissertation*, Newtonian science is true of the sensible world, to which sensibility gives us access; and the understanding grasps principles of divine and moral perfection in a distinct intelligible world, which are paradigms for measuring everything in the sensible world. So on this view our knowledge of the intelligible world is a priori because it does not depend on sensibility, and this a priori knowledge furnishes principles for judging the sensible world because in some way the sensible world itself conforms to or imitates the intelligible world.

Soon after writing the *Inaugural Dissertation*, however, Kant expressed doubts about this view. As he explained in a February 21, 1772 letter to his friend and former student, Marcus Herz:

In my dissertation I was content to explain the nature of intellectual representations in a merely negative way, namely, to state that they were not modifications of the soul brought about by the object. However, I silently passed over the further question of how a representation that refers to an object without being in any way affected by it can be possible.... [B]y what means are these [intellectual representations] given to us, if not by the way in which they affect us? And if such intellectual representations depend on our inner activity, whence comes the agreement that they are supposed to have with objects – objects that are nevertheless not possibly produced thereby?...[A]s to how my understanding may form for itself concepts of things completely a priori, with which concepts the things must necessarily agree, and as to how my understanding may formulate real principles concerning the possibility of such concepts, with which principles experience must be in exact agreement and which nevertheless are independent of experience – this question, of how the faculty of understanding achieves this conformity with the things themselves, is still left in a state of obscurity. (10:130–131)

Here Kant entertains doubts about how a priori knowledge of an intelligible world would be possible. The position of the *Inaugural Dissertation* is that the intelligible world is independent of the human understanding and of the sensible world, both of which (in different ways) conform to the intelligible world. But, leaving aside questions about what it means for the sensible world to conform to an intelligible world, how is it possible for the human understanding to conform to or grasp an intelligible world? If the

intelligible world is independent of our understanding, then it seems that we could grasp it only if we are passively affected by it in some way. But for Kant sensibility is our passive or receptive capacity to be affected by objects that are independent of us (2:392, A51/B75). So the only way we could grasp an intelligible world that is independent of us is through sensibility, which means that our knowledge of it could not be a priori. The pure understanding alone could at best enable us to form representations of an intelligible world. But since these intellectual representations would entirely “depend on our inner activity,” as Kant says to Herz, we have no good reason to believe that they would conform to an independent intelligible world. Such a priori intellectual representations could well be figments of the brain that do not correspond to anything independent of the human mind. In any case, it is completely mysterious how there might come to be a correspondence between purely intellectual representations and an independent intelligible world.

Kant’s strategy in the *Critique* is similar to that of the *Inaugural Dissertation* in that both works attempt to reconcile modern science with traditional morality and religion by relegating them to distinct sensible and intelligible worlds, respectively. But the *Critique* gives a far more modest and yet revolutionary account of a priori knowledge. As Kant’s letter to Herz suggests, the main problem with his view in the *Inaugural Dissertation* is that it tries to explain the possibility of a priori knowledge about a world that is entirely independent of the human mind. This turned out to be a dead end, and Kant never again maintained that we can have a priori knowledge about an intelligible world precisely because such a world would be entirely independent of us. However, Kant’s revolutionary position in the *Critique* is that we can have a priori knowledge about the general structure of the sensible world because it is not entirely independent of the human mind. The sensible world, or the world of appearances, is constructed by the human mind from a combination of sensory matter that we receive passively and a priori forms that are supplied by our cognitive faculties. We can have a priori knowledge only about aspects of the sensible world that reflect the a priori forms supplied by our cognitive faculties. In Kant’s words, “we can cognize of things a priori only what we ourselves have put into them” (Bxviii). So according to the *Critique*, a priori knowledge is possible only if and to the extent that the sensible world itself depends on the way the human mind structures its experience.

Kant characterizes this new constructivist view of experience in the *Critique* through an analogy with the revolution wrought by Copernicus in astronomy:

Up to now it has been assumed that all our cognition must conform to the objects; but all attempts to find out something about them a priori

through concepts that would extend our cognition have, on this presupposition, come to nothing. Hence let us once try whether we do not get farther with the problems of metaphysics by assuming that the objects must conform to our cognition, which would agree better with the requested possibility of an a priori cognition of them, which is to establish something about objects before they are given to us. This would be just like the first thoughts of Copernicus, who, when he did not make good progress in the explanation of the celestial motions if he assumed that the entire celestial host revolves around the observer, tried to see if he might not have greater success if he made the observer revolve and left the stars at rest. Now in metaphysics we can try in a similar way regarding the intuition of objects. If intuition has to conform to the constitution of the objects, then I do not see how we can know anything of them a priori; but if the object (as an object of the senses) conforms to the constitution of our faculty of intuition, then I can very well represent this possibility to myself. Yet because I cannot stop with these intuitions, if they are to become cognitions, but must refer them as representations to something as their object and determine this object through them, I can assume either that the concepts through which I bring about this determination also conform to the objects, and then I am once again in the same difficulty about how I could know anything about them a priori, or else I assume that the objects, or what is the same thing, the experience in which alone they can be cognized (as given objects) conforms to those concepts, in which case I immediately see an easier way out of the difficulty, since experience itself is a kind of cognition requiring the understanding, whose rule I have to presuppose in myself before any object is given to me, hence a priori, which rule is expressed in concepts a priori, to which all objects of experience must therefore necessarily conform, and with which they must agree. (Bxvi-xviii)

As this passage suggests, what Kant has changed in the *Critique* is primarily his view about the role and powers of the understanding, since he already held in the *Inaugural Dissertation* that sensibility contributes the forms of space and time – which he calls pure (or a priori) intuitions (2:397) – to our cognition of the sensible world. But the *Critique* claims that pure understanding too, rather than giving us insight into an intelligible world, is limited to providing

forms – which he calls pure or a priori concepts – that structure our cognition of the sensible world. So now both sensibility and understanding work together to construct cognition of the sensible world, which therefore conforms to the a priori forms that are supplied by our cognitive faculties: the a priori intuitions of sensibility and the a priori concepts of the understanding. This account is analogous to the heliocentric revolution of Copernicus in astronomy because both require contributions from the observer to be factored into explanations of phenomena, although neither reduces phenomena to the contributions of observers alone.[6] The way celestial phenomena appear to us on earth, according to Copernicus, is affected by both the motions of celestial bodies and the motion of the earth, which is not a stationary body around which everything else revolves. For Kant, analogously, the phenomena of human experience depend on both the sensory data that we receive passively through sensibility and the way our mind actively processes this data according to its own a priori rules. These rules supply the general framework in which the sensible world and all the objects (or phenomena) in it appear to us. So the sensible world and its phenomena are not entirely independent of the human mind, which contributes its basic structure.

How does Kant's Copernican revolution in philosophy improve on the strategy of the *Inaugural Dissertation* for reconciling modern science with traditional morality and religion? First, it gives Kant a new and ingenious way of placing modern science on an a priori foundation. He is now in a position to argue that we can have a priori knowledge about the basic laws of modern science because those laws reflect the human mind's contribution to structuring our experience. In other words, the sensible world necessarily conforms to certain fundamental laws – such as that every event has a cause – because the human mind constructs it according to those laws. Moreover, we can identify those laws by reflecting on the conditions of possible experience, which reveals that it would be impossible for us to experience a world in which, for example, any given event fails to have a cause. From this Kant concludes that metaphysics is indeed possible in the sense that we can have a priori knowledge that the entire sensible world – not just our actual experience, but any possible human experience – necessarily conforms to certain laws. Kant calls this immanent metaphysics or the metaphysics of experience, because it deals with the essential principles that are immanent to human experience.

But, second, if “we can cognize of things a priori only what we ourselves have put into them,” then we cannot have a priori knowledge about things whose existence and nature are entirely independent of the human mind, which Kant calls things in themselves (Bxviii). In his words: “[F]rom this deduction of our faculty of cognizing a priori [...] there emerges a very strange result [...], namely that with this faculty we can never get beyond the boundaries of possible experience, [...and] that such cognition

reaches appearances only, leaving the thing in itself as something actual for itself but uncognized by us" (Bxix-xx). That is, Kant's constructivist foundation for scientific knowledge restricts science to the realm of appearances and implies that transcendent metaphysics – i.e., a priori knowledge of things in themselves that transcend possible human experience – is impossible. In the *Critique* Kant thus rejects the insight into an intelligible world that he defended in the *Inaugural Dissertation*, and he now claims that rejecting knowledge about things in themselves is necessary for reconciling science with traditional morality and religion. This is because he claims that belief in God, freedom, and immortality have a strictly moral basis, and yet adopting these beliefs on moral grounds would be unjustified if we could know that they were false. "Thus," Kant says, "I had to deny knowledge in order to make room for faith" (Bxxx). Restricting knowledge to appearances and relegating God and the soul to an unknowable realm of things in themselves guarantees that it is impossible to disprove claims about God and the freedom or immortality of the soul, which moral arguments may therefore justify us in believing. Moreover, the determinism of modern science no longer threatens the freedom required by traditional morality, because science and therefore determinism apply only to appearances, and there is room for freedom in the realm of things in themselves, where the self or soul is located. We cannot know (theoretically) that we are free, because we cannot know anything about things in themselves. But there are especially strong moral grounds for the belief in human freedom, which acts as "the keystone" supporting other morally grounded beliefs (5:3-4). In this way, Kant replaces transcendent metaphysics with a new practical science that he calls the metaphysics of morals. It thus turns out that two kinds of metaphysics are possible: the metaphysics of experience (or nature) and the metaphysics of morals, both of which depend on Kant's Copernican revolution in philosophy.

3. Transcendental idealism

Perhaps the central and most controversial thesis of the *Critique of Pure Reason* is that human beings experience only appearances, not things in themselves; and that space and time are only subjective forms of human intuition that would not subsist in themselves if one were to abstract from all subjective conditions of human intuition. Kant calls this thesis transcendental idealism.

[7] One of his best summaries of it is arguably the following:

We have therefore wanted to say that all our intuition is nothing but the representation of appearance; that the things that we intuit are not in themselves what we intuit them to be, nor are their relations so constituted in themselves as they appear to us; and that if we remove our own subject or even only the subjective constitution of the

senses in general, then all constitution, all relations of objects in space and time, indeed space and time themselves would disappear, and as appearances they cannot exist in themselves, but only in us. What may be the case with objects in themselves and abstracted from all this receptivity of our sensibility remains entirely unknown to us. We are acquainted with nothing except our way of perceiving them, which is peculiar to us, and which therefore does not necessarily pertain to every being, though to be sure it pertains to every human being. We are concerned solely with this. Space and time are its pure forms, sensation in general its matter. We can cognize only the former a priori, i.e., prior to all actual perception, and they are therefore called pure intuition; the latter, however, is that in our cognition that is responsible for its being called a posteriori cognition, i.e., empirical intuition. The former adheres to our sensibility absolutely necessarily, whatever sort of sensations we may have; the latter can be very different. (A42/B59–60) [8]

Kant introduces transcendental idealism in the part of the *Critique* called the Transcendental Aesthetic, and scholars generally agree that for Kant transcendental idealism encompasses at least the following claims:

- In some sense, human beings experience only appearances, not things in themselves.
- Space and time are not things in themselves, or determinations of things in themselves that would remain if one abstracted from all subjective conditions of human intuition. [Kant labels this conclusion a) at A26/B42 and again at A32–33/B49. It is at least a crucial part of what he means by calling space and time transcendently ideal (A28/B44, A35–36/B52)].
- Space and time are nothing other than the subjective forms of human sensible intuition. [Kant labels this conclusion b) at A26/B42 and again at A33/B49–50].
- Space and time are empirically real, which means that “everything that can come before us externally as an object” is in both space and time, and that our internal intuitions of ourselves are in time (A28/B44, A34–35/B51–51).

But scholars disagree widely on how to interpret these claims, and there is no such thing as the standard interpretation of Kant’s transcendental idealism. Two general types of interpretation have been especially influential, however. This section provides an overview of these two interpretations, although it should be

emphasized that much important scholarship on transcendental idealism does not fall neatly into either of these two camps.

3.1 The two-objects interpretation

The two-objects reading is the traditional interpretation of Kant's transcendental idealism. It goes back to the earliest review of the *Critique* – the so-called Göttingen review by Christian Garve (1742–1798) and J. G. Feder (1740–1821)[[9](#)] – and it was the dominant way of interpreting Kant's transcendental idealism during his own lifetime. It has been a live interpretive option since then and remains so today, although it no longer enjoys the dominance that it once did.[[10](#)]

According to the two-objects interpretation, transcendental idealism is essentially a metaphysical thesis that distinguishes between two classes of objects: appearances and things in themselves. Another name for this view is the two-worlds interpretation, since it can also be expressed by saying that transcendental idealism essentially distinguishes between a world of appearances and another world of things in themselves.

Things in themselves, on this interpretation, are absolutely real in the sense that they would exist and have whatever properties they have even if no human beings were around to perceive them. Appearances, on the other hand, are not absolutely real in that sense, because their existence and properties depend on human perceivers. Moreover, whenever appearances do exist, in some sense they exist in the mind of human perceivers. So appearances are mental entities or mental representations. This, coupled with the claim that we experience only appearances, makes transcendental idealism a form of phenomenalism on this interpretation, because it reduces the objects of experience to mental representations. All of our experiences – all of our perceptions of objects and events in space, even those objects and events themselves, and all non-spatial but still temporal thoughts and feelings – fall into the class of appearances that exist in the mind of human perceivers. These appearances cut us off entirely from the reality of things in themselves, which are non-spatial and non-temporal. Yet Kant's theory, on this interpretation, nevertheless requires that things in themselves exist, because they must transmit to us the sensory data from which we construct appearances. In principle we cannot know how things in themselves affect our senses, because our experience and knowledge is limited to the world of appearances constructed by and in the mind. Things in themselves are therefore a sort of theoretical posit, whose existence and role are required by the theory but are not directly verifiable.

The main problems with the two-objects interpretation are philosophical. Most readers of Kant who have interpreted his transcendental idealism in this way have been – often very – critical of it, for reasons such as the following:

First, at best Kant is walking a fine line in claiming on the one hand that we can have no knowledge about things in themselves, but on the other hand that we know that things in themselves exist, that they affect our senses, and that they are non-spatial and non-temporal. At worst his theory depends on contradictory claims about what we can and cannot know about things in themselves. This objection was influentially articulated by Jacobi, when he complained that “without that presupposition [of things in themselves] I could not enter into the system, but with it I could not stay within it” (Jacobi 1787, 336).

Second, even if that problem is surmounted, it has seemed to many that Kant’s theory, interpreted in this way, implies a radical form of skepticism that traps each of us within the contents of our own mind and cuts us off from reality. Some versions of this objection proceed from premises that Kant rejects. One version maintains that things in themselves are real while appearances are not, and hence that on Kant’s view we cannot have experience or knowledge of reality. But Kant denies that appearances are unreal: they are just as real as things in themselves but are in a different metaphysical class. Another version claims that truth always involves a correspondence between mental representations and things in themselves, from which it would follow that on Kant’s view it is impossible for us to have true beliefs about the world. But just as Kant denies that things in themselves are the only (or privileged) reality, he also denies that correspondence with things in themselves is the only kind of truth. Empirical judgments are true just in case they correspond with their empirical objects in accordance with the a priori principles that structure all possible human experience. But the fact that Kant can appeal in this way to an objective criterion of empirical truth that is internal to our experience has not been enough to convince some critics that Kant is innocent of an unacceptable form of skepticism, mainly because of his insistence on our irreparable ignorance about things in themselves.

Third and finally, Kant’s denial that things in themselves are spatial or temporal has struck many of his readers as incoherent. The role of things in themselves, on the two-object interpretation, is to affect our senses and thereby to provide the sensory data from which our cognitive faculties construct appearances within the framework of our a priori intuitions of space and time and a priori concepts such as causality. But if there is no space, time, change, or causation in the realm of things in themselves, then how can things in themselves affect us? Transcendental affection seems to involve a causal relation between things in themselves and our sensibility. If this is simply the way we unavoidably think about transcendental affection, because we can give positive content to this thought only by employing the concept of a cause, while it is nevertheless strictly false that things in themselves affect us causally, then it seems not only that we are ignorant of how things in themselves

really affect us. It seems, rather, to be incoherent that things in themselves could affect us at all if they are not in space or time.

3.2 The two-aspects interpretation

The two-aspects reading attempts to interpret Kant's transcendental idealism in a way that enables it to be defended against at least some of these objections. On this view, transcendental idealism does not distinguish between two classes of objects but rather between two different aspects of one and the same class of objects. For this reason it is also called the one-world interpretation, since it holds that there is only one world in Kant's ontology, and that at least some objects in that world have two different aspects: one aspect that appears to us, and another aspect that does not appear to us. That is, appearances are aspects of the same objects that also exist in themselves. So, on this reading, appearances are not mental representations, and transcendental idealism is not a form of phenomenalism.[\[11\]](#)

There are at least two main versions of the two-aspects theory. One version treats transcendental idealism as a metaphysical theory according to which objects have two aspects in the sense that they have two sets of properties: one set of relational properties that appear to us and are spatial and temporal, and another set of intrinsic properties that do not appear to us and are not spatial or temporal (Langton 1998). This property-dualist interpretation faces epistemological objections similar to those faced by the two-objects interpretation, because we are in no better position to acquire knowledge about properties that do not appear to us than we are to acquire knowledge about objects that do not appear to us. Moreover, this interpretation also seems to imply that things in themselves are spatial and temporal, since appearances have spatial and temporal properties, and on this view appearances are the same objects as things in themselves. But Kant explicitly denies that space and time are properties of things in themselves.

A second version of the two-aspects theory departs more radically from the traditional two-objects interpretation by denying that transcendental idealism is at bottom a metaphysical theory. Instead, it interprets transcendental idealism as a fundamentally epistemological theory that distinguishes between two standpoints on the objects of experience: the human standpoint, from which objects are viewed relative to epistemic conditions that are peculiar to human cognitive faculties (namely, the a priori forms of our sensible intuition); and the standpoint of an intuitive intellect, from which the same objects could be known in themselves and independently of any epistemic conditions (Allison 2004). Human beings cannot really take up the latter standpoint but can form only an empty concept of things as they exist in themselves by abstracting from all the content of our experience and leaving only the purely formal thought of an object in general. So transcendental idealism, on this interpretation, is essentially the

thesis that we are limited to the human standpoint, and the concept of a thing in itself plays the role of enabling us to chart the boundaries of the human standpoint by stepping beyond them in abstract (but empty) thought.

One criticism of this epistemological version of the two-aspects theory is that it avoids the objections to other interpretations by attributing to Kant a more limited project than the text of the *Critique* warrants. There are passages that support this reading.[12] But there are also many passages in both editions of the *Critique* in which Kant describes appearances as representations in the mind and in which his distinction between appearances and things in themselves is given not only epistemological but metaphysical significance.[13] It is unclear whether all of these texts admit of a single, consistent interpretation.

4. The transcendental deduction

The transcendental deduction is the central argument of the *Critique of Pure Reason* and one of the most complex and difficult texts in the history of philosophy. Given its complexity, there are naturally many different ways of interpreting the deduction.[14] This brief overview provides one perspective on some of its main ideas.

The transcendental deduction occurs in the part of the *Critique* called the Analytic of Concepts, which deals with the a priori concepts that, on Kant's view, our understanding uses to construct experience together with the a priori forms of our sensible intuition (space and time), which he discussed in the Transcendental Aesthetic. Kant calls these a priori concepts "categories," and he argues elsewhere (in the so-called metaphysical deduction) that they include such concepts as substance and cause. The goal of the transcendental deduction is to show that we have a priori concepts or categories that are objectively valid, or that apply necessarily to all objects in the world that we experience. To show this, Kant argues that the categories are necessary conditions of experience, or that we could not have experience without the categories. In Kant's words:

[T]he objective validity of the categories, as a priori concepts, rests on the fact that through them alone is experience possible (as far as the form of thinking is concerned). For they then are related necessarily and a priori to objects of experience, since only by means of them can any object of experience be thought at all.

The transcendental deduction of all a priori concepts therefore has a principle toward which the entire investigation must be directed, namely this: that they

must be recognized as a priori conditions of the possibility of experiences (whether of the intuition that is encountered in them, or of the thinking). Concepts that supply the objective ground of the possibility of experience are necessary just for that reason. (A93-94/B126)

The strategy Kant employs to argue that the categories are conditions of experience is the main source of both the obscurity and the ingenuity of the transcendental deduction. His strategy is to argue that the categories are necessary specifically for self-consciousness, for which Kant often uses the Leibnizian term “apperception.”

4.1 Self-consciousness

One way to approach Kant’s argument is to contrast his view of self-consciousness with two alternative views that he rejects. Each of these views, both Kant’s and those he rejects, can be seen as offering competing answers the question: what is the source of our sense of an ongoing and invariable self that persists throughout all the changes in our experience?

The first answer to this question that Kant rejects is that self-consciousness arises from some particular content being present in each of one’s representations. This material conception of self-consciousness, as we may call it, is suggested by Locke’s account of personal identity. According to Locke, “it being the same consciousness that makes a Man be himself to himself, personal Identity depends on that only, whether it be annexed only to one individual Substance, or can be continued in a succession of several Substances” (*Essay* 2.27.10). What Locke calls “the same consciousness” may be understood as some representational content that is always present in my experience and that both identifies any experience as mine and gives me a sense of a continuous self by virtue of its continual presence in my experience. One problem with this view, Kant believes, is that there is no such representational content that is invariably present in experience, so the sense of an ongoing self cannot possibly arise from that non-existent content (what Locke calls “consciousness”) being present in each of one’s representations. In Kant’s words, self-consciousness “does not yet come about by my accompanying each representation with consciousness, but rather by my adding one representation to the other and being conscious of their synthesis. Therefore it is only because I can combine a manifold of given representations in one consciousness that it is possible for me to represent the identity of the consciousness in these representations” (B133). Here Kant claims, against the Lockean view, that self-consciousness arises from combining (or synthesizing) representations with one another regardless of their content. In short, Kant has a formal conception of self-consciousness rather than a material one. Since no particular

content of my experience is invariable, self-consciousness must derive from my experience having an invariable form or structure, and consciousness of the identity of myself through all of my changing experiences must consist in awareness of the formal unity and law-governed regularity of my experience. The continuous form of my experience is the necessary correlate for my sense of a continuous self.

There are at least two possible versions of the formal conception of self-consciousness: a realist and an idealist version. On the realist version, nature itself is law-governed and we become self-conscious by attending to its law-governed regularities, which also makes this an empiricist view of self-consciousness. The idea of an identical self that persists throughout all of our experience, on this view, arises from the law-governed regularity of nature, and our representations exhibit order and regularity because reality itself is ordered and regular. Kant rejects this realist view and embraces a conception of self-consciousness that is both formal and idealist. According to Kant, the formal structure of our experience, its unity and law-governed regularity, is an achievement of our cognitive faculties rather than a property of reality in itself. Our experience has a constant form because our mind constructs experience in a law-governed way. So self-consciousness, for Kant, consists in awareness of the mind's law-governed activity of synthesizing or combining sensible data to construct a unified experience. As he expresses it, "this unity of consciousness would be impossible if in the cognition of the manifold the mind could not become conscious of the identity of the function by means of which this manifold is synthetically combined into one cognition" (A108).

Kant argues for this formal idealist conception of self-consciousness, and against the formal realist view, on the grounds that "we can represent nothing as combined in the object without having previously combined it ourselves" (B130). In other words, even if reality in itself were law-governed, its laws could not simply migrate over to our mind or imprint themselves on us while our mind is entirely passive. We must exercise an active capacity to represent the world as combined or ordered in a law-governed way, because otherwise we could not represent the world as law-governed even if it were law-governed in itself. Moreover, this capacity to represent the world as law-governed must be a priori because it is a condition of self-consciousness, and we would already have to be self-conscious in order to learn from our experience that there are law-governed regularities in the world. So it is necessary for self-consciousness that we exercise an a priori capacity to represent the world as law-governed. But this would also be sufficient for self-consciousness if we could exercise our a priori capacity to represent the world as law-governed even if reality in itself were not law-governed. In that case, the realist and empiricist conception of self-consciousness would be false, and the formal idealist view would be true.

Kant's confidence that no empiricist account could possibly explain self-consciousness may be based on his assumption that the sense of self each of us has, the thought of oneself as identical throughout all of one's changing experiences, involves necessity and universality, which on his view are the hallmarks of the a priori. This assumption is reflected in what we may call Kant's principle of apperception: "The **I think** must **be able** to accompany all my representations; for otherwise something would be represented in me that could not be thought at all, which is as much as to say that the representation would either be impossible or else at least would be nothing for me" (B131-132).^[15] Notice the claims about necessity and universality embodied in the words "must" and "all" here. Kant is saying that for a representation to count as mine, it must necessarily be accessible to conscious awareness in some (perhaps indirect) way: I must be able to accompany it with "I think...." All of my representations must be accessible to consciousness in this way (but they need not actually be conscious), because again that is simply what makes a representation count as mine. Self-consciousness for Kant therefore involves a priori knowledge about the necessary and universal truth expressed in this principle of apperception, and a priori knowledge cannot be based on experience.

Kant may have developed this thread of his argument in the transcendental deduction after reading Johann Nicolaus Tetens (1736-1807) rather than through a direct encounter with Locke's texts (Tetens 1777, Kitcher 2011). On the subject of self-consciousness, Tetens was a follower of Locke and also engaged with Hume's arguments for rejecting a continuing self. So Kant's actual opponents in the deduction may have been Lockean and Humean positions as represented by Tetens, as well as rationalist views that Kant would have encountered directly in texts by Leibniz, Wolff, and some of their followers.

4.2 Objectivity and judgment

On the basis of this formal idealist conception of self-consciousness, Kant's argument (at least one central thread of it) moves through two more conditions of self-consciousness in order to establish the objective validity of the categories. The next condition is that self-consciousness requires me to represent an objective world distinct from my subjective representations – that is, distinct from my thoughts about and sensations of that objective world. Kant uses this connection between self-consciousness and objectivity to insert the categories into his argument.

In order to be self-conscious, I cannot be wholly absorbed in the contents of my perceptions but must distinguish myself from the rest of the world. But if self-consciousness is an achievement of the mind, then how does the mind achieve this sense that there is a distinction between the I that perceives and the contents of its perceptions? According to Kant, the mind achieves this sense by

distinguishing representations that necessarily belong together from representations that are not necessarily connected but are merely associated in a contingent way. Consider Kant's example of the perception of a house (B162). Imagine a house that is too large to fit into your visual field from your vantage point near its front door. Now imagine that you walk around the house, successively perceiving each of its sides. Eventually you perceive the entire house, but not all at once, and you judge that each of your representations of the sides of the house necessarily belong together (as sides of one house) and that anyone who denied this would be mistaken. But now imagine that you grew up in this house and associate a feeling of nostalgia with it. You would not judge that representations of this house are necessarily connected with feelings of nostalgia. That is, you would not think that other people seeing the house for the first time would be mistaken if they denied that it is connected with nostalgia, because you recognize that this house is connected with nostalgia for you but not necessarily for everyone. Yet you distinguish this merely subjective connection from the objective connection between sides of the house, which is objective because the sides of the house necessarily belong together "in the object," because this connection holds for everyone universally, and because it is possible to be mistaken about it. The point here is not that we must successfully identify which representations necessarily belong together and which are merely associated contingently, but rather that to be self-conscious we must at least make this general distinction between objective and merely subjective connections of representations.

At this point (at least in the second edition text) Kant introduces the key claim that judgment is what enables us to distinguish objective connections of representations that necessarily belong together from merely subjective and contingent associations: "[A] judgment is nothing other than the way to bring given cognitions to the objective unity of apperception. That is the aim of the copula **is** in them: to distinguish the objective unity of given representations from the subjective. For this word designates the relation of the representations to the original apperception and its necessary unity" (B141-142). Kant is speaking here about the mental act of judging that results in the formation of a judgment. Judging is an act of what Kant calls synthesis, which he defines as "the action of putting different representations together with each other and comprehending their manifoldness in one cognition" (A77/B103). In other words, to synthesize is in general to combine several representations into a single (more) complex representation, and to judge is specifically to combine concepts into a judgment – that is, to join a subject concept to a predicate concept by means of the copula, as in "the body is heavy" or "the house is four-sided." Judgments need not be true, of course, but they always have a truth value (true or false) because they make claims to objective validity. When I say, by contrast, that "If I carry

a body, I feel a pressure of weight,” or that “if I see this house, I feel nostalgia,” I am not making a judgment about the object (the body or the house) but rather I am expressing a subjective association that may apply only to me (B142).[16]

Kant’s reference to the necessary unity of apperception or self-consciousness in the quotation above means (at least) that the action of judging is the way our mind achieves self-consciousness. We must represent an objective world in order to distinguish ourselves from it, and we represent an objective world by judging that some representations necessarily belong together. Moreover, recall from 4.1 that, for Kant, we must have an a priori capacity to represent the world as law-governed, because “we can represent nothing as combined (or connected) in the object without having previously combined it ourselves” (B130). It follows that objective connections in the world cannot simply imprint themselves on our mind. Rather, experience of an objective world must be constructed by exercising an a priori capacity to judge, which Kant calls the faculty of understanding (A80–81/B106). The understanding constructs experience by providing the a priori rules, or the framework of necessary laws, in accordance with which we judge representations to be objective. These rules are the pure concepts of the understanding or categories, which are therefore conditions of self-consciousness, since they are rules for judging about an objective world, and self-consciousness requires that we distinguish ourselves from an objective world.

Kant identifies the categories in what he calls the metaphysical deduction, which precedes the transcendental deduction.[17] Very briefly, since the categories are a priori rules for judging, Kant argues that an exhaustive table of categories can be derived from a table of the basic logical forms of judgments. For example, according to Kant the logical form of the judgment that “the body is heavy” would be singular, affirmative, categorical, and assertoric. But since categories are not mere logical functions but instead are rules for making judgments about objects or an objective world, Kant arrives at his table of categories by considering how each logical function would structure judgments about objects (within our spatio-temporal forms of intuition). For example, he claims that categorical judgments express a logical relation between subject and predicate that corresponds to the ontological relation between substance and accident; and the logical form of a hypothetical judgment expresses a relation that corresponds to cause and effect. Taken together with this argument, then, the transcendental deduction argues that we become self-conscious by representing an objective world of substances that interact according to causal laws.

4.3 The law-giver of nature

The final condition of self-consciousness that Kant adds to the preceding conditions is that our understanding must cooperate

with sensibility to construct one, unbounded, and unified space-time to which all of our representations may be related.

To see why this further condition is required, consider that so far we have seen why Kant holds that we must represent an objective world in order to be self-conscious, but we could represent an objective world even if it were not possible to relate all of our representations to this objective world. For all that has been said so far, we might still have unruly representations that we cannot relate in any way to the objective framework of our experience. On Kant's view, this would be a problem because, as we have seen, he holds that self-consciousness involves universality and necessity: according to his principle of apperception, "the **I think** must **be able** to accompany all my representations" (B131). Yet if, on the one hand, I had representations that I could not relate in some way to an objective world, then I could not accompany those representations with "I think" or recognize them as my representations, because I can say "I think..." about any given representation only by relating it to an objective world, according to the argument just discussed. So I must be able to relate any given representation to an objective world in order for it to count as mine. On the other hand, self-consciousness would also be impossible if I represented multiple objective worlds, even if I could relate all of my representations to some objective world or other. In that case, I could not become conscious of an identical self that has, say, representation 1 in space-time A and representation 2 in space-time B. It may be possible to imagine disjointed spaces and times, but it is not possible to represent them as objectively real. So self-consciousness requires that I can relate all of my representations to a single objective world.

The reason why I must represent this one objective world by means of a unified and unbounded space-time is that, as Kant argued in the *Transcendental Aesthetic*, space and time are the pure forms of human intuition. If we had different forms of intuition, then our experience would still have to constitute a unified whole in order for us to be self-conscious, but this would not be a spatio-temporal whole. Given that space and time are our forms of intuition, however, our understanding must still cooperate with sensibility to construct a spatio-temporal whole of experience because, once again, "we can represent nothing as combined in the object without having previously combined it ourselves," and "all combination [...] is an action of the understanding" (B130). So Kant distinguishes between space and time as pure forms of intuition, which belong solely to sensibility; and the formal intuitions of space and time (or space-time), which are unified by the understanding (B160-161). These formal intuitions are the spatio-temporal whole within which our understanding constructs experience in accordance with the categories.[18]

The most important implication of Kant's claim that the understanding constructs a single whole of experience to which all of our representations can be related is that, since he defines nature "regarded materially" as "the sum total of all appearances" and he has argued that the categories are objectively valid of all possible appearances, on his view it follows that our categories are the source of the fundamental laws of nature "regarded formally" (B163, 165). So Kant concludes on this basis that the understanding is the true law-giver of nature. In his words: "all appearances in nature, as far as their combination is concerned, stand under the categories, on which nature (considered merely as nature in general) depends, as the original ground of its necessary lawfulness (as nature regarded formally)" (B165). Or more strongly: "we ourselves bring into the appearances that order and regularity that we call **nature**, and moreover we would not be able to find it there if we, or the nature of our mind, had not originally put it there. [...] The understanding is thus not merely a faculty for making rules through the comparison of the appearances: it is itself the legislation for nature, i.e., without understanding there would not be any nature at all" (A125-126).

5. Morality and freedom

Having examined two central parts of Kant's positive project in theoretical philosophy from the *Critique of Pure Reason*, transcendental idealism and the transcendental deduction, let us now turn to his practical philosophy in the *Critique of Practical Reason*. Since Kant's philosophy is deeply systematic, this section begins with a preliminary look at how his theoretical and practical philosophy fit together (see also section 7).

5.1 Theoretical and practical autonomy

The fundamental idea of Kant's philosophy is human autonomy. So far we have seen this in Kant's constructivist view of experience, according to which our understanding is the source of the general laws of nature. "Autonomy" literally means giving the law to oneself, and on Kant's view our understanding provides laws that constitute the a priori framework of our experience. Our understanding does not provide the matter or content of our experience, but it does provide the basic formal structure within which we experience any matter received through our senses. Kant's central argument for this view is the transcendental deduction, according to which it is a condition of self-consciousness that our understanding constructs experience in this way. So we may call self-consciousness the highest principle of Kant's theoretical philosophy, since it is (at least) the basis for all of our a priori knowledge about the structure of nature.

Kant's moral philosophy is also based on the idea of autonomy. He holds that there is a single fundamental principle of morality, on which all specific moral duties are based. He calls this moral law

(as it is manifested to us) the categorical imperative (see [5.4](#)). The moral law is a product of reason, for Kant, while the basic laws of nature are products of our understanding. There are important differences between the senses in which we are autonomous in constructing our experience and in morality. For example, Kant regards understanding and reason as different cognitive faculties, although he sometimes uses “reason” in a wide sense to cover both. [\[19\]](#) The categories and therefore the laws of nature are dependent on our specifically human forms of intuition, while reason is not. The moral law does not depend on any qualities that are peculiar to human nature but only on the nature of reason as such, although its manifestation to us as a categorical imperative (as a law of duty) reflects the fact that the human will is not necessarily determined by pure reason but is also influenced by other incentives rooted in our needs and inclinations; and our specific duties deriving from the categorical imperative do reflect human nature and the contingencies of human life. Despite these differences, however, Kant holds that we give the moral law to ourselves, as we also give the general laws of nature to ourselves, though in a different sense. Moreover, we each necessarily give the same moral law to ourselves, just as we each construct our experience in accordance with the same categories. To summarize:

- Theoretical philosophy is about how the world is (A633/B661). Its highest principle is self-consciousness, on which our knowledge of the basic laws of nature is based. Given sensory data, our understanding constructs experience according to these a priori laws.
- Practical philosophy is about how the world ought to be (ibid., A800–801/B828–829). Its highest principle is the moral law, from which we derive duties that command how we ought to act in specific situations. Kant also claims that reflection on our moral duties and our need for happiness leads to the thought of an ideal world, which he calls the highest good (see section [6](#)). Given how the world is (theoretical philosophy) and how it ought to be (practical philosophy), we aim to make the world better by constructing or realizing the highest good.

So both parts of Kant’s philosophy are about autonomously constructing a world, but in different senses. In theoretical philosophy, we use our categories and forms of intuition to construct a world of experience or nature. In practical philosophy, we use the moral law to construct the idea of a moral world or a realm of ends that guides our conduct (4:433), and ultimately to transform the natural world into the highest good. Finally, transcendental idealism is the framework within which these two parts of Kant’s philosophy fit together (20:311). Theoretical philosophy deals with appearances, to which our knowledge is strictly limited; and practical philosophy deals with things in

themselves, although it does not give us knowledge about things in themselves but only provides rational justification for certain beliefs about them for practical purposes.

To understand Kant's arguments that practical philosophy justifies certain beliefs about things in themselves, it is necessary to see them in the context of his criticism of German rationalist metaphysics. The three traditional topics of Leibniz-Wolffian special metaphysics were rational psychology, rational cosmology, and rational theology, which dealt, respectively, with the human soul, the world-whole, and God. In the part of the *Critique of Pure Reason* called the Transcendental Dialectic, Kant argues against the Leibniz-Wolffian view that human beings are capable of a priori knowledge in each of these domains, and he claims that the errors of Leibniz-Wolffian metaphysics are due to an illusion that has its seat in the nature of human reason itself. According to Kant, human reason necessarily produces ideas of the soul, the world-whole, and God; and these ideas unavoidably produce the illusion that we have a priori knowledge about transcendent objects corresponding to them. This is an illusion, however, because in fact we are not capable of a priori knowledge about any such transcendent objects. Nevertheless, Kant attempts to show that these illusory ideas have a positive, practical use. He thus reframes Leibniz-Wolffian special metaphysics as a practical science that he calls the metaphysics of morals. On Kant's view, our ideas of the soul, the world-whole, and God provide the content of morally justified beliefs about human immortality, human freedom, and the existence of God, respectively; but they are not proper objects of speculative knowledge.[\[20\]](#)

5.2 Freedom

The most important belief about things in themselves that Kant thinks only practical philosophy can justify concerns human freedom. Freedom is important because, on Kant's view, moral appraisal presupposes that we are free in the sense that we have the ability to do otherwise. To see why, consider Kant's example of a man who commits a theft (5:95ff.). Kant holds that in order for this man's action to be morally wrong, it must have been within his control in the sense that it was within his power at the time not to have committed the theft. If this was not within his control at the time, then, while it may be useful to punish him in order to shape his behavior or to influence others, it nevertheless would not be correct to say that his action was morally wrong. Moral rightness and wrongness apply only to free agents who control their actions and have it in their power, at the time of their actions, either to act rightly or not. According to Kant, this is just common sense.

On these grounds, Kant rejects a type of compatibilism that he calls the "comparative concept of freedom" and associates with Leibniz (5:96–97). (Note that Kant has a specific type of compatibilism in mind, which I will refer to simply as "compatibilism," although

there may be other types of compatibilism that do not fit Kant's characterization of that view). On the compatibilist view, as Kant understands it, I am free whenever the cause of my action is within me. So I am unfree only when something external to me pushes or moves me, but I am free whenever the proximate cause of my body's movement is internal to me as an "acting being" (5:96). If we distinguish between involuntary convulsions and voluntary bodily movements, then on this view free actions are just voluntary bodily movements. Kant ridicules this view as a "wretched subterfuge" that tries to solve an ancient philosophical problem "with a little quibbling about words" (ibid.). This view, he says, assimilates human freedom to "the freedom of a turnspit," or a projectile in flight, or the motion of a clock's hands (5:96-97). The proximate causes of these movements are internal to the turnspit, the projectile, and the clock at the time of the movement. This cannot be sufficient for moral responsibility.

Why not? The reason, Kant says, is ultimately that the causes of these movements occur in time. Return to the theft example. A compatibilist would say that the thief's action is free because its proximate cause is inside him, and because the theft was not an involuntary convulsion but a voluntary action. The thief decided to commit the theft, and his action flowed from this decision. According to Kant, however, if the thief's decision is a natural phenomenon that occurs in time, then it must be the effect of some cause that occurred in a previous time. This is an essential part of Kant's Newtonian worldview and is grounded in the a priori laws (specifically, the category of cause and effect) in accordance with which our understanding constructs experience: every event has a cause that begins in an earlier time. If that cause too was an event occurring in time, then it must also have a cause beginning in a still earlier time, etc. All natural events occur in time and are thoroughly determined by causal chains that stretch backwards into the distant past. So there is no room for freedom in nature, which is deterministic in a strong sense.

The root of the problem, for Kant, is time. Again, if the thief's choice to commit the theft is a natural event in time, then it is the effect of a causal chain extending into the distant past. But the past is out of his control now, in the present. Once the past is past, he can't change it. On Kant's view, that is why his actions would not be in his control in the present if they are determined by events in the past. Even if he could control those past events in the past, he cannot control them now. But in fact past events were not in his control in the past either if they too were determined by events in the more distant past, because eventually the causal antecedents of his action stretch back before his birth, and obviously events that occurred before his birth were never in his control. So if the thief's choice to commit the theft is a natural event in time, then it is not now and never was in his control, and he could not have done

otherwise than to commit the theft. In that case, it would be a mistake to hold him morally responsible for it.

Compatibilism, as Kant understands it, therefore locates the issue in the wrong place. Even if the cause of my action is internal to me, if it is in the past – for example, if my action today is determined by a decision I made yesterday, or from the character I developed in childhood – then it is not within my control now. The real issue is not whether the cause of my action is internal or external to me, but whether it is in my control now. For Kant, however, the cause of my action can be within my control now only if it is not in time. This is why Kant thinks that transcendental idealism is the only way to make sense of the kind of freedom that morality requires. Transcendental idealism allows that the cause of my action may be a thing in itself outside of time: namely, my noumenal self, which is free because it is not part of nature. No matter what kind of character I have developed or what external influences act on me, on Kant's view all of my intentional, voluntary actions are immediate effects of my noumenal self, which is causally undetermined (5:97–98). My noumenal self is an uncaused cause outside of time, which therefore is not subject to the deterministic laws of nature in accordance with which our understanding constructs experience.

Many puzzles arise on this picture that Kant does not resolve. For example, if my understanding constructs all appearances in my experience of nature, not only appearances of my own actions, then why am I responsible only for my own actions but not for everything that happens in the natural world? Moreover, if I am not alone in the world but there are many noumenal selves acting freely and incorporating their free actions into the experience they construct, then how do multiple transcendently free agents interact? How do you integrate my free actions into the experience that your understanding constructs?[21] In spite of these unsolved puzzles, Kant holds that we can make sense of moral appraisal and responsibility only by thinking about human freedom in this way, because it is the only way to prevent natural necessity from undermining both.

Finally, since Kant invokes transcendental idealism to make sense of freedom, interpreting his thinking about freedom leads us back to disputes between the two-objects and two-aspects interpretations of transcendental idealism. On the face of it, the two-objects interpretation seems to make better sense of Kant's view of transcendental freedom than the two-aspects interpretation. If morality requires that I am transcendently free, then it seems that my true self, and not just an aspect of my self, must be outside of time, according to Kant's argument. But applying the two-objects interpretation to freedom raises problems of its own, since it involves making a distinction between noumenal and phenomenal selves that does not arise on the two-aspects view.

If only my noumenal self is free, and freedom is required for moral responsibility, then my phenomenal self is not morally responsible. But how are my noumenal and phenomenal selves related, and why is punishment inflicted on phenomenal selves? It is unclear whether and to what extent appealing to Kant's theory of freedom can help to settle disputes about the proper interpretation of transcendental idealism, since there are serious questions about the coherence of Kant's theory on either interpretation.

5.3 The fact of reason

Can we know that we are free in this transcendental sense? Kant's response is tricky. On the one hand, he distinguishes between theoretical knowledge and morally justified belief (A820-831/B848-859). We do not have theoretical knowledge that we are free or about anything beyond the limits of possible experience, but we are morally justified in believing that we are free in this sense. On the other hand, Kant also uses stronger language than this when discussing freedom. For example, he says that "among all the ideas of speculative reason freedom is the only one the possibility of which we know a priori, though without having any insight into it, because it is the condition of the moral law, which we do know." In a footnote to this passage, Kant explains that we know freedom a priori because "were there no freedom, the moral law would not be encountered at all in ourselves," and on Kant's view everyone does encounter the moral law a priori (5:4). For this reason, Kant claims that the moral law "proves" the objective, "though only practical, undoubted reality" of freedom (5:48-49). So Kant wants to say that we do have knowledge of the reality of freedom, but that this is practical knowledge of a practical reality, or cognition "only for practical purposes," by which he means to distinguish it from theoretical knowledge based on experience or reflection on the conditions of experience (5:133). Our practical knowledge of freedom is based instead on the moral law. The difference between Kant's stronger and weaker language seems mainly to be that his stronger language emphasizes that our belief or practical knowledge about freedom is unshakeable and that it in turn provides support for other morally grounded beliefs in God and the immortality of the soul.

Kant calls our consciousness of the moral law, our awareness that the moral law binds us or has authority over us, the "fact of reason" (5:31-32, 42-43, 47, 55). So, on his view, the fact of reason is the practical basis for our belief or practical knowledge that we are free. Kant insists that this moral consciousness is "undeniable," "a priori," and "unavoidable" (5:32, 47, 55). Every human being has a conscience, a common sense grasp of morality, and a firm conviction that he or she is morally accountable. We may have different beliefs about the source of morality's authority - God, social convention, human reason. We may arrive at different conclusions about what morality requires in specific situations. And

we may violate our own sense of duty. But we all have a conscience, and an unshakeable belief that morality applies to us. According to Kant, this belief cannot and does not need to be justified or “proved by any deduction” (5:47). It is just a ground-level fact about human beings that we hold ourselves morally accountable. But Kant is making a normative claim here as well: it is also a fact, which cannot and does not need to be justified, that we are morally accountable, that morality does have authority over us. Kant holds that philosophy should be in the business of defending this common sense moral belief, and that in any case we could never prove or disprove it (4:459).

Kant may hold that the fact of reason, or our consciousness of moral obligation, implies that we are free on the grounds that ought implies can. In other words, Kant may believe that it follows from the fact that we ought (morally) to do something that we can or are able to do it. This is suggested, for example, by a passage in which Kant asks us to imagine someone threatened by his prince with immediate execution unless he “give[s] false testimony against an honorable man whom the prince would like to destroy under a plausible pretext.” Kant says that “[h]e would perhaps not venture to assert whether he would do it or not, but he must admit without hesitation that it would be possible for him. He judges, therefore, that he can do something because he is aware that he ought to do it and cognizes freedom within him, which, without the moral law, would have remained unknown to him” (5:30). This is a hypothetical example of an action not yet carried out. It seems that pangs of guilt about the immorality of an action that you carried out in the past, on this reasoning, would imply more directly that you have (or at least had) the ability to act otherwise than you did, and therefore that you are free in Kant’s sense.

5.4 The categorical imperative

In both the *Groundwork of the Metaphysics of Morals* and the *Critique of Practical Reason*, Kant also gives a more detailed argument for the conclusion that morality and freedom reciprocally imply one another, which is sometimes called the reciprocity thesis (Allison 1990). On this view, to act morally is to exercise freedom, and the only way to fully exercise freedom is to act morally. Kant’s arguments for this view differ in these texts, but the general structure of his argument in the *Critique of Practical Reason* may be summarized as follows.

First, it follows from the basic idea of having a will that to act at all is to act on some principle, or what Kant calls a maxim. A maxim is a subjective rule or policy of action: it says what you are doing and why. Kant gives as examples the maxims “to let no insult pass unavenged” and “to increase my wealth by every safe means” (5:19, 27). We may be unaware of our maxims, we may not act consistently on the same maxims, and our maxims may not be consistent with one another. But Kant holds that since we are

rational beings our actions always aim at some sort of end or goal, which our maxim expresses. The goal of an action may be something as basic as gratifying a desire, or it may be something more complex such as becoming a doctor or a lawyer. In any case, the causes of our actions are never our desires or impulses, on Kant's view. If I act to gratify some desire, then I choose to act on a maxim that specifies the gratification of that desire as the goal of my action. For example, if I desire some coffee, then I may act on the maxim to go to a cafe and buy some coffee in order to gratify that desire.

Second, Kant distinguishes between two basic kinds of principles or rules that we can act on: what he calls material and formal principles. To act in order to satisfy some desire, as when I act on the maxim to go for coffee at a cafe, is to act on a material principle (5:21ff.). Here the desire (for coffee) fixes the goal, which Kant calls the object or matter of the action, and the principle says how to achieve that goal (go to a cafe). Corresponding to material principles, on Kant's view, are what he calls hypothetical imperatives. A hypothetical imperative is a principle of rationality that says I should act in a certain way if I choose to satisfy some desire. If maxims in general are rules that describe how one does act, then imperatives in general prescribe how one should act. An imperative is hypothetical if it says how I should act only if I choose to pursue some goal in order to gratify a desire (5:20). This, for example, is a hypothetical imperative: if you want coffee, then go to the cafe. This hypothetical imperative applies to you only if you desire coffee and choose to gratify that desire.

In contrast to material principles, formal principles describe how one acts without making reference to any desires. This is easiest to understand through the corresponding kind of imperative, which Kant calls a categorical imperative. A categorical imperative commands unconditionally that I should act in some way. So while hypothetical imperatives apply to me only on the condition that I have and set the goal of satisfying the desires that they tell me how to satisfy, categorical imperatives apply to me no matter what my goals and desires may be. Kant regards moral laws as categorical imperatives, which apply to everyone unconditionally. For example, the moral requirement to help others in need does not apply to me only if I desire to help others in need, and the duty not to steal is not suspended if I have some desire that I could satisfy by stealing. Moral laws do not have such conditions but rather apply unconditionally. That is why they apply to everyone in the same way.

Third, insofar as I act only on material principles or hypothetical imperatives, I do not act freely, but rather I act only to satisfy some desire(s) that I have, and what I desire is not ultimately within my control. To some limited extent we are capable of rationally shaping our desires, but insofar as we choose to act in order to

satisfy desires we are choosing to let nature govern us rather than governing ourselves (5:118). We are always free in the sense that we always have the capacity to govern ourselves rationally instead of letting our desires set our ends for us. But we may (freely) fail to exercise that capacity. Moreover, since Kant holds that desires never cause us to act, but rather we always choose to act on a maxim even when that maxim specifies the satisfaction of a desire as the goal of our action, it also follows that we are always free in the sense that we freely choose our maxims. Nevertheless, our actions are not free in the sense of being autonomous if we choose to act only on material principles, because in that case we do not give the law to ourselves, but instead we choose to allow nature in us (our desires) to determine the law for our actions.

Finally, the only way to act freely in the full sense of exercising autonomy is therefore to act on formal principles or categorical imperatives, which is also to act morally. Kant does not mean that acting autonomously requires that we take no account of our desires, which would be impossible (5:25, 61). Rather, he holds that we typically formulate maxims with a view to satisfying our desires, but that “as soon as we draw up maxims of the will for ourselves” we become immediately conscious of the moral law (5:29). This immediate consciousness of the moral law takes the following form:

I have, for example, made it my maxim to increase my wealth by every safe means. Now I have a deposit in my hands, the owner of which has died and left no record of it. This is, naturally, a case for my maxim. Now I want only to know whether that maxim could also hold as a universal practical law. I therefore apply the maxim to the present case and ask whether it could indeed take the form of a law, and consequently whether I could through my maxim at the same time give such a law as this: that everyone may deny a deposit which no one can prove has been made. I at once become aware that such a principle, as a law, would annihilate itself since it would bring it about that there would be no deposits at all. (5:27)

In other words, to assess the moral permissibility of my maxim, I ask whether everyone could act on it, or whether it could be willed as a universal law. The issue is not whether it would be good if everyone acted on my maxim, or whether I would like it, but only whether it would be possible for my maxim to be willed as a universal law. This gets at the form, not the matter or content, of the maxim. A maxim has morally permissible form, for Kant, only if it could be willed as a universal law. If my maxim fails this test, as this one does, then it is morally impermissible for me to act on it.

If my maxim passes the universal law test, then it is morally permissible for me to act on it, but I fully exercise my autonomy only if my fundamental reason for acting on this maxim is that it is morally permissible or required that I do so. Imagine that I am moved by a feeling of sympathy to formulate the maxim to help someone in need. In this case, my original reason for formulating this maxim is that a certain feeling moved me. Such feelings are not entirely within my control and may not be present when someone actually needs my help. But this maxim passes Kant's test: it could be willed as a universal law that everyone help others in need from motives of sympathy. So it would not be wrong to act on this maxim when the feeling of sympathy so moves me. But helping others in need would not fully exercise my autonomy unless my fundamental reason for doing so is not that I have some feeling or desire, but rather that it would be right or at least permissible to do so. Only when such a purely formal principle supplies the fundamental motive for my action do I act autonomously.

So the moral law is a law of autonomy in the sense that "freedom and unconditional practical law reciprocally imply each another" (5:29). Even when my maxims are originally suggested by my feelings and desires, if I act only on morally permissible (or required) maxims because they are morally permissible (or required), then my actions will be autonomous. And the reverse is true as well: for Kant this is the only way to act autonomously.[\[22\]](#)

6. The highest good and practical postulates

Kant holds that reason unavoidably produces not only consciousness of the moral law but also the idea of a world in which there is both complete virtue and complete happiness, which he calls the highest good. Our duty to promote the highest good, on Kant's view, is the sum of all moral duties, and we can fulfill this duty only if we believe that the highest good is a possible state of affairs. Furthermore, we can believe that the highest good is possible only if we also believe in the immortality of the soul and the existence of God, according to Kant. On this basis, he claims that it is morally necessary to believe in the immortality of the soul and the existence of God, which he calls postulates of pure practical reason. This section briefly outlines Kant's view of the highest good and his argument for these practical postulates in the *Critique of Practical Reason* and other works.

6.1 The highest good

In the previous section we saw that, on Kant's view, the moral law is a purely formal principle that commands us to act only on maxims that have what he calls lawgiving form, which maxims have only if they can be willed as universal laws. Moreover, our fundamental reason for choosing to act on such maxims should be that they have this lawgiving form, rather than that acting on them would achieve some end or goal that would satisfy a desire (5:27).

For example, I should help others in need not, at bottom, because doing so would make me feel good, even if it would, but rather because it is right; and it is right (or permissible) to help others in need because this maxim can be willed as a universal law.

Although Kant holds that the morality of an action depends on the form of its maxim rather than its end or goal, he nevertheless claims both that every human action has an end and that we are unavoidably concerned with the consequences of our actions (4:437; 5:34; 6:5–7, 385). This is not a moral requirement but simply part of what it means to be a rational being. Moreover, Kant also holds the stronger view that it is an unavoidable feature of human reason that we form ideas not only about the immediate and near-term consequences of our actions, but also about ultimate consequences. This is the practical manifestation of reason's general demand for what Kant calls "the unconditioned" (5:107–108).^[23] In particular, since we naturally have desires and inclinations, and our reason has "a commission" to attend to the satisfaction of our desires and inclinations, on Kant's view we unavoidably form an idea of the maximal satisfaction of all our inclinations and desires, which he calls happiness (5:61, 22, 124). This idea is indeterminate, however, since nobody can know "what he really wishes and wills" and thus what would make him completely happy (4:418). We also form the idea of a moral world or realm of ends, in which everyone acts only in accordance with maxims that can be universal laws (A808/B836, 4:433ff.).

But neither of these ideas by itself expresses our unconditionally complete end, as human reason demands in its practical use. A perfectly moral world by itself would not constitute our "whole and complete good [...]" even in the judgment of an impartial reason," because it is human nature also to need happiness (5:110, 25). And happiness by itself would not be unconditionally good, because moral virtue is a condition of worthiness to be happy (5:111). So our unconditionally complete end must combine both virtue and happiness. In Kant's words, "virtue and happiness together constitute possession of the highest good in a person, and happiness distributed in exact proportion to morality (as the worth of a person and his worthiness to be happy) constitutes the *highest good* of a possible world" (5:110–111). It is this ideal world combining complete virtue with complete happiness that Kant normally has in mind when he discusses the highest good.

Kant says that we have a duty to promote the highest good, taken in this sense (5:125). He does not mean, however, to be identifying some new duty that is not derived from the moral law, in addition to all the particular duties we have that are derived from the moral law.^[24] For example, he is not claiming that in addition to my duties to help others in need, not to commit theft, etc., I also have the additional duty to represent the highest good as the final end of all moral conduct, combined with happiness, and to promote that

end. Rather, as we have seen, Kant holds that it is an unavoidable feature of human reasoning, instead of a moral requirement, that we represent all particular duties as leading toward the promotion of the highest good. So the duty to promote the highest good is not a particular duty at all, but the sum of all our duties derived from the moral law – it “does not increase the number of morality’s duties but rather provides these with a special point of reference for the unification of all ends” (6:5). Nor does Kant mean that anyone has a duty to realize or actually bring about the highest good through their own power, although his language sometimes suggests this (5:113, 122). Rather, at least in his later works Kant claims that only the common striving of an entire “ethical community” can actually produce the highest good, and that the duty of individuals is to promote (but not single-handedly produce) this end with all of their strength by doing what the moral law commands (6:97–98, 390–394).[25]

Finally, according to Kant we must conceive of the highest good as a possible state of affairs in order to fulfill our duty to promote it. Here Kant does not mean that we unavoidably represent the highest good as possible, since his view is that we must represent it as possible only if we are to fulfill our duty of promoting it, and yet we may fail at doing our duty. Rather, we have a choice about whether to conceive of the highest good as possible, to regard it as impossible, or to remain noncommittal (5:144–145). But we can fulfill our duty of promoting the highest good only by choosing to conceive of the highest good as possible, because we cannot promote any end without believing that it is possible to achieve that end (5:122). So fulfilling the sum of all moral duties to promote the highest good requires believing that a world of complete virtue and happiness is not simply “a phantom of the mind” but could actually be realized (5:472).

6.2 The postulates of pure practical reason

Kant argues that we can comply with our duty to promote the highest good only if we believe in the immortality of the soul and the existence of God. This is because to comply with that duty we must believe that the highest good is possible, and yet to believe that the highest good is possible we must believe that the soul is immortal and that God exists, according to Kant.[26]

Consider first Kant’s moral argument for belief in immortality. The highest good, as we have seen, would be a world of complete morality and happiness. But Kant holds that it is impossible for “a rational being of the sensible world” to exhibit “complete conformity of dispositions with the moral law,” which he calls “holiness,” because we can never extirpate the propensity of our reason to give priority to the incentives of inclination over the incentive of duty, which propensity Kant calls radical evil (5:122, 6:37). Kant claims that the moral law nevertheless requires holiness, however, and that it therefore “can only be found in an

endless progress toward that complete conformity,” or progress that goes to infinity (5:122). This does not mean that we can substitute endless progress toward complete conformity with the moral law for holiness in the concept of the highest good, but rather that we must represent that complete conformity as an infinite progress toward the limit of holiness. Kant continues: “This endless progress is, however, possible only on the presupposition of the existence and personality of the same rational being continuing endlessly (which is called the immortality of the soul). Hence the highest good is practically possible only on the presupposition of the immortality of the soul, so that this, as inseparable with the moral law, is a postulate of pure practical reason” (ibid.). Kant’s idea is not that we should imagine ourselves attaining holiness later although we are not capable of it in this life. Rather, his view is that we must represent holiness as continual progress toward complete conformity of our dispositions with the moral law that begins in this life and extends into infinity.

Kant’s moral argument for belief in God in the *Critique of Practical Reason* may be summarized as follows. Kant holds that virtue and happiness are not just combined but necessarily combined in the idea of the highest good, because only possessing virtue makes one worthy of happiness – a claim that Kant seems to regard as part of the content of the moral law (4:393; 5:110, 124). But we can represent virtue and happiness as necessarily combined only by representing virtue as the efficient cause of happiness. This means that we must represent the highest good not simply as a state of affairs in which everyone is both happy and virtuous, but rather as one in which everyone is happy because they are virtuous (5:113–114, 124). However, it is beyond the power of human beings, both individually and collectively, to guarantee that happiness results from virtue, and we do not know any law of nature that guarantees this either. Therefore, we must conclude that the highest good is impossible, unless we postulate “the existence of a cause of nature, distinct from nature, which contains the ground of this connection, namely the exact correspondence of happiness with morality” (5:125). This cause of nature would have to be God since it must have both understanding and will. Kant probably does not conceive of God as the efficient cause of a happiness that is rewarded in a future life to those who are virtuous in this one. Rather, his view is probably that we represent our endless progress toward holiness, beginning with this life and extending into infinity, as the efficient cause of our happiness, which likewise begins in this life and extends to a future one, in accordance with teleological laws that God authors and causes to harmonize with efficient causes in nature (A809–812/B837–840; 5:127–131, 447–450).

Both of these arguments are subjective in the sense that, rather than attempting to show how the world must be constituted objectively in order for the highest good to be possible, they purport to show only how we must conceive of the highest good in

order to be subjectively capable both of representing it as possible and of fulfilling our duty to promote it. But Kant also claims that both arguments have an objective basis: first, in the sense that it cannot be proven objectively either that immortality or God's existence are impossible; and, second, in the sense that both arguments proceed from a duty to promote the highest good that is based not on the subjective character of human reason but on the moral law, which is objectively valid for all rational beings. So while it is not, strictly speaking, a duty to believe in God or immortality, we must believe both in order to fulfill our duty to promote the highest good, given the subjective character of human reason.

To see why, consider what would happen if we did not believe in God or immortality, according to Kant. In the *Critique of Pure Reason*, Kant seems to say that this would leave us without any incentive to be moral, and even that the moral law would be invalid without God and immortality (A813/B841, A468/B496). But Kant later rejects this view (8:139). His mature view is that our reason would be in conflict with itself if we did not believe in God and immortality, because pure practical reason would represent the moral law as authoritative for us and so present us with an incentive that is sufficient to determine our will; but pure theoretical (i.e., speculative) reason would undermine this incentive by declaring morality an empty ideal, since it would not be able to conceive of the highest good as possible (5:121, 143, 471–472, 450–453). In other words, the moral law would remain valid and provide any rational being with sufficient incentive to act from duty, but we would be incapable of acting as rational beings, since “it is a condition of having reason at all [...] that its principles and affirmations must not contradict one another” (5:120). The only way to bring speculative and practical reason “into that relation of equality in which reason in general can be used purposively” is to affirm the postulates on the grounds that pure practical reason has primacy over speculative reason. This means, Kant explains, that if the capacity of speculative reason “does not extend to establishing certain propositions affirmatively, although they do not contradict it, as soon as these same propositions belong inseparably to the practical interest of pure reason it must accept them [...] being mindful, however, that these are not its insights but are yet extensions of its use from another, namely a practical perspective” (5:121). The primacy of practical reason is a key element of Kant's response to the crisis of the Enlightenment, since he holds that reason deserves the sovereign authority entrusted to it by the Enlightenment only on this basis.

7. The unity of nature and freedom

This final section briefly discusses how Kant attempts to unify the theoretical and practical parts of his philosophical system in the *Critique of the Power of Judgment*.

7.1 The great chasm

In the Preface and Introduction to the *Critique of the Power of Judgment*, Kant announces that his goal in the work is to “bring [his] entire critical enterprise to an end” by bridging the “gulf” or “chasm” that separates the domain of his theoretical philosophy (discussed mainly in the *Critique of Pure Reason*) from the domain of his practical philosophy (discussed mainly in the *Critique of Practical Reason*) (5:170, 176, 195). In his words: “The understanding legislates a priori for nature, as object of the senses, for a theoretical cognition of it in a possible experience. Reason legislates a priori for freedom and its own causality, as the supersensible in the subject, for an unconditioned practical cognition. The domain of the concept of nature under the one legislation and that of the concept of freedom under the other are entirely barred from any mutual influence that they could have on each other by themselves (each in accordance with its fundamental laws) by the great chasm that separates the supersensible from the appearances” (5:195).

One way to understand the problem Kant is articulating here is to consider it once again in terms of the crisis of the Enlightenment. [27] The crisis was that modern science threatened to undermine traditional moral and religious beliefs, and Kant’s response is to argue that in fact these essential interests of humanity are consistent with one another when reason is granted sovereignty and practical reason is given primacy over speculative reason. But the transcendental idealist framework within which Kant develops this response seems to purchase the consistency of these interests at the price of sacrificing a unified view of the world and our place in it. If science applies only to appearances, while moral and religious beliefs refer to things in themselves or “the supersensible,” then how can we integrate these into a single conception of the world that enables us to transition from the one domain to the other? Kant’s solution is to introduce a third a priori cognitive faculty, which he calls the reflecting power of judgment, that gives us a teleological perspective on the world. Reflecting judgment provides the concept of teleology or purposiveness that bridges the chasm between nature and freedom, and thus unifies the theoretical and practical parts of Kant’s philosophy into a single system (5:196–197).

It is important to Kant that a third faculty independent of both understanding and reason provides this mediating perspective, because he holds that we do not have adequate theoretical grounds for attributing objective teleology to nature itself, and yet regarding nature as teleological solely on moral grounds would only heighten the disconnect between our scientific and moral ways of viewing the world. Theoretical grounds do not justify us in attributing objective teleology to nature, because it is not a condition of self-consciousness that our understanding construct

experience in accordance with the concept of teleology, which is not among Kant's categories or the principles of pure understanding that ground the fundamental laws of nature. That is why his theoretical philosophy licenses us only in attributing mechanical causation to nature itself. To this limited extent, Kant is sympathetic to the dominant strain in modern philosophy that banishes final causes from nature and instead treats nature as nothing but matter in motion, which can be fully described mathematically. But Kant wants somehow to reconcile this mechanistic view of nature with a conception of human agency that is essentially teleological. As we saw in the previous section, Kant holds that every human action has an end and that the sum of all moral duties is to promote the highest good. It is essential to Kant's approach, however, to maintain the autonomy of both understanding (in nature) and reason (in morality), without allowing either to encroach on the other's domain, and yet to harmonize them in a single system. This harmony can be orchestrated only from an independent standpoint, from which we do not judge how nature is constituted objectively (that is the job of understanding) or how the world ought to be (the job of reason), but from which we merely regulate or reflect on our cognition in a way that enables us to regard it as systematically unified. According to Kant, this is the task of reflecting judgment, whose a priori principle is to regard nature as purposive or teleological, "but only as a regulative principle of the faculty of cognition" (5:197).

7.2 The purposiveness of nature

In the *Critique of the Power of Judgment*, Kant discusses four main ways in which reflecting judgment leads us to regard nature as purposive: first, it leads us to regard nature as governed by a system of empirical laws; second, it enables us to make aesthetic judgments; third, it leads us to think of organisms as objectively purposive; and, fourth, it ultimately leads us to think about the final end of nature as a whole.[\[28\]](#)

First, reflecting judgment enables us to discover empirical laws of nature by leading us to regard nature as if it were the product of intelligent design (5:179–186). We do not need reflecting judgment to grasp the a priori laws of nature based on our categories, such as that every event has a cause. But in addition to these a priori laws nature is also governed by particular, empirical laws, such as that fire causes smoke, which we cannot know without consulting experience. To discover these laws, we must form hypotheses and devise experiments on the assumption that nature is governed by empirical laws that we can grasp (Bxiii–xiv). Reflecting judgment makes this assumption through its principle to regard nature as purposive for our understanding, which leads us to treat nature as if its empirical laws were designed to be understood by us (5:180–181). Since this principle only regulates our cognition but is not

constitutive of nature itself, this does not amount to assuming that nature really is the product of intelligent design, which according to Kant we are not justified in believing on theoretical grounds. Rather, it amounts only to approaching nature in the practice of science as if it were designed to be understood by us. We are justified in doing this because it enables us to discover empirical laws of nature. But it is only a regulative principle of reflecting judgment, not genuine theoretical knowledge, that nature is purposive in this way.

Second, Kant thinks that aesthetic judgments about both beauty and sublimity involve a kind of purposiveness, and that the beauty of nature in particular suggests to us that nature is hospitable to our ends. According to his aesthetic theory, we judge objects to be beautiful not because they gratify our desires, since aesthetic judgments are disinterested, but rather because apprehending their form stimulates what he calls the harmonious “free play” of our understanding and imagination, in which we take a distinctively aesthetic pleasure (5:204–207, 217–218, 287). So beauty is not a property of objects, but a relation between their form and the way our cognitive faculties work. Yet we make aesthetic judgments that claim intersubjective validity because we assume that there is a common sense that enables all human beings to communicate aesthetic feeling (5:237–240, 293–296). Beautiful art is intentionally created to stimulate this universally communicable aesthetic pleasure, although it is effective only when it seems unintentional (5:305–307). Natural beauty, however, is unintentional: landscapes do not know how to stimulate the free play of our cognitive faculties, and they do not have the goal of giving us aesthetic pleasure. In both cases, then, beautiful objects appear purposive to us because they give us aesthetic pleasure in the free play of our faculties, but they also do not appear purposive because they either do not or do not seem to do this intentionally. Kant calls this relation between our cognitive faculties and the formal qualities of objects that we judge to be beautiful “subjective purposiveness” (5:221). Although it is only subjective, the purposiveness exhibited by natural beauty in particular may be interpreted as a sign that nature is hospitable to our moral interests (5:300). Moreover, Kant also interprets the experience of sublimity in nature as involving purposiveness. But in this case it is not so much the purposiveness of nature as our own purpose or “vocation” as moral beings that we become aware of in the experience of the sublime, in which the size and power of nature stand in vivid contrast to the superior power of our reason (5:257–260, 267–269).

Third, Kant argues that reflecting judgment enables us to regard living organisms as objectively purposive, but only as a regulative principle that compensates for our inability to fully understand them mechanistically, which reflects the limitations of our cognitive faculties rather than any intrinsic teleology in nature. We cannot

fully understand organisms mechanistically because they are “self-organizing” beings, whose parts are “combined into a whole by being reciprocally the cause and effect of their form” (5:373–374). The parts of a watch are also possible only through their relation to the whole, but that is because the watch is designed and produced by some rational being. An organism, by contrast, produces and sustains itself, which is inexplicable to us unless we attribute to organisms purposes by analogy with human art (5:374–376). But Kant claims that it is only a regulative principle of reflecting judgment to regard organisms in this way, and that we are not justified in attributing objective purposiveness to organisms themselves, since it is only “because of the peculiar constitution of my cognitive faculties [that] I cannot judge about the possibility of those things and their generation except by thinking of a cause for these acts in accordance with intentions” (5:397–398). Specifically, we cannot understand how a whole can be the cause of its own parts because we depend on sensible intuition for the content of our thoughts and therefore must think the particular (intuition) first by subsuming it under the general (a concept). To see that this is just a limitation of the human, discursive intellect, imagine a being with an intuitive understanding whose thought does not depend, as ours does, on receiving sensory information passively, but rather creates the content of its thought in the act of thinking it. Such a (divine) being could understand how a whole can be the cause of its parts, since it could grasp a whole immediately without first thinking particulars and then combining them into a whole (5:401–410). Therefore, since we have a discursive intellect and cannot know how things would appear to a being with an intuitive intellect, and yet we can only think of organisms teleologically, which excludes mechanism, Kant now says that we must think of both mechanism and teleology only as regulative principles that we need to explain nature, rather than as constitutive principles that describe how nature is intrinsically constituted (5:410ff.).

Fourth, Kant concludes the *Critique of the Power of Judgment* with a long appendix arguing that reflecting judgment supports morality by leading us to think about the final end of nature, which we can only understand in moral terms, and that conversely morality reinforces a teleological conception of nature. Once it is granted on theoretical grounds that we must understand certain parts of nature (organisms) teleologically, although only as a regulative principle of reflecting judgment, Kant says we may go further and regard the whole of nature as a teleological system (5:380–381). But we can regard the whole of nature as a teleological system only by employing the idea of God, again only regulatively, as its intelligent designer. This involves attributing what Kant calls external purposiveness to nature – that is, attributing purposes to God in creating nature (5:425). What, then, is God’s final end in creating nature? According to Kant, the final end of nature must be human beings, but only as moral beings (5:435, 444–445). This is

because only human beings use reason to set and pursue ends, using the rest of nature as means to their ends (5:426-427). Moreover, Kant claims that human happiness cannot be the final end of nature, because as we have seen he holds that happiness is not unconditionally valuable (5:430-431). Rather, human life has value not because of what we passively enjoy, but only because of what we actively do (5:434). We can be fully active and autonomous, however, only by acting morally, which implies that God created the world so that human beings could exercise moral autonomy. Since we also need happiness, this too may be admitted as a conditioned and consequent end, so that reflecting judgment eventually leads us to the highest good (5:436). But reflection on conditions of the possibility of the highest good leads again to Kant's moral argument for belief in God's existence, which in turn reinforces the teleological perspective on nature with which reflecting judgment began.

Thus Kant argues that although theoretical and practical philosophy proceed from separate and irreducible starting points – self-consciousness as the highest principle for our cognition of nature, and the moral law as the basis for our knowledge of freedom – reflecting judgment unifies them into a single, teleological worldview that assigns preeminent value to human autonomy.